Project page

AdaRadar: Rate Adaptive Spectral Compression for Radar-based Perception

Jinho Park[1], Se Young Chun[2], Mingoo Seok[1]

[1]Columbia University

[2]Seoul National University

jp4327@columbia.edu, sychun@snu.ac.kr, ms4415@columbia.edu

Table of Contents

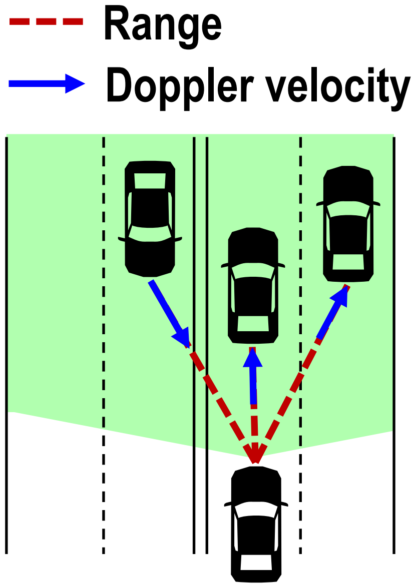
- Motivation
- Proposed methods
  - Spectral pruning
  - Adaptive rate feedback
- Experimental Results
  - Qualitative results
  - Rate-accuracy tradeoff
  - Compression performance against baseline work
- Conclusion

Table of Contents

Radar Complements Camera & LiDAR

Vision is vulnerable to low-light and occlusion

LiDAR can be affected by scattering from rain & snow

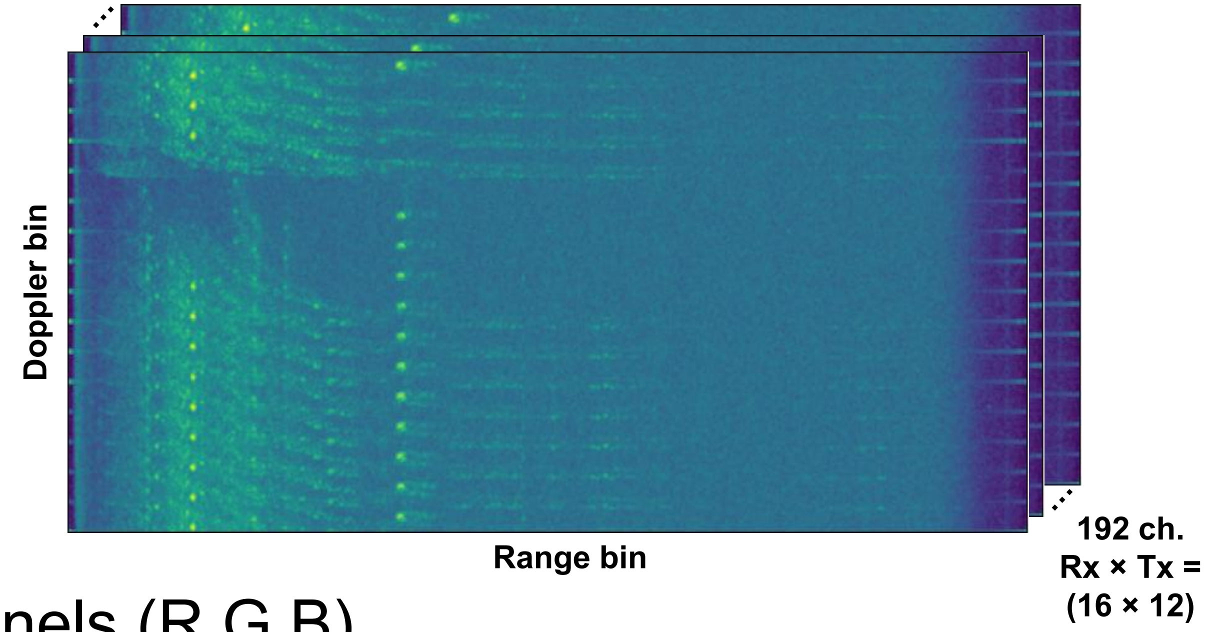


- Radar-based perception is a critical sensing modality in autonomous driving pipeline as it complements camera and LiDAR.
- Radar is robust to lighting and weather conditions.
- Its ability to measure relative range and Doppler velocity.

Radar Poses Bottleneck on Link BW

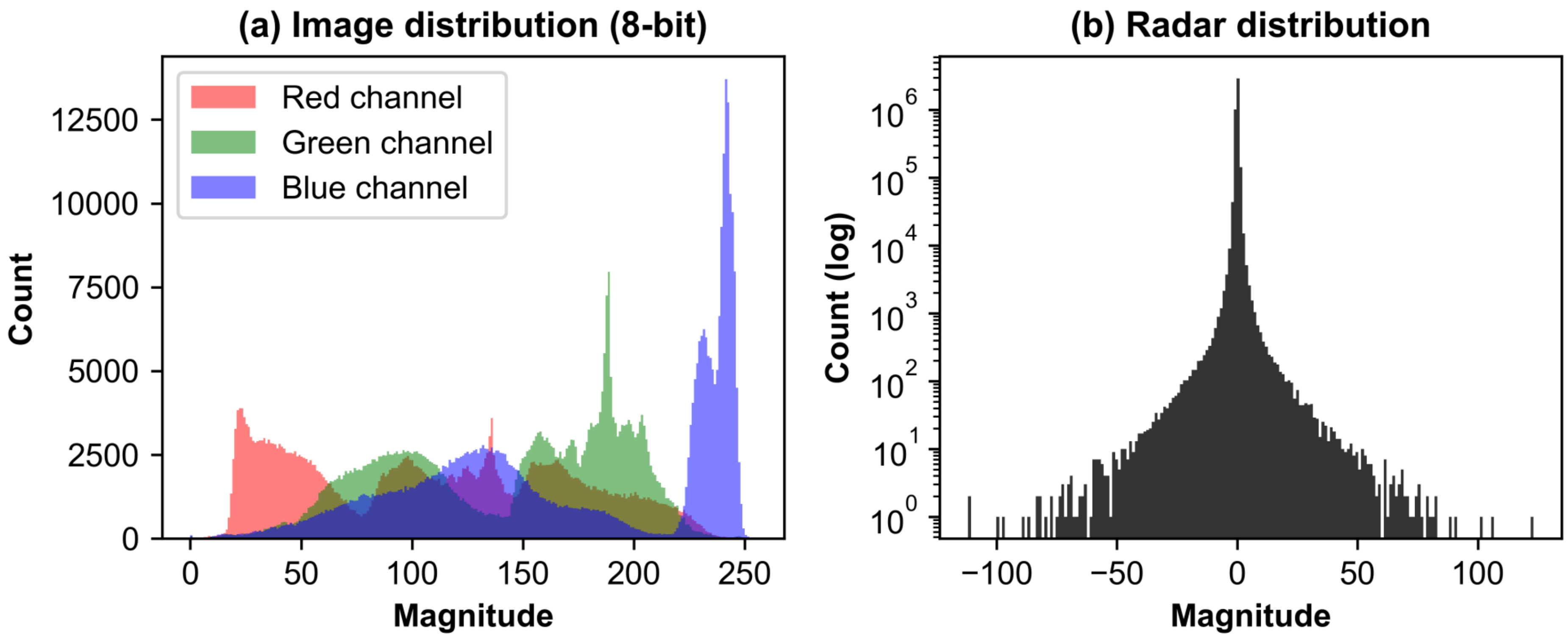
Image from KITTI

Radar from RADIal



- Unlike a camera, which only has 3 channels (R,G,B).
- Raw radar tensor has quadratically scaling channels to receiver (Rx) and transmitter (Tx) antennas.
- e.g., a single frame from a 16 x 12 cascaded MIMO radar system, configured with 512 range bins and 256 Doppler bins, amounts to roughly 100 MB.

Absence of Radar CODECs



a) Natural images span a broad dynamic range.

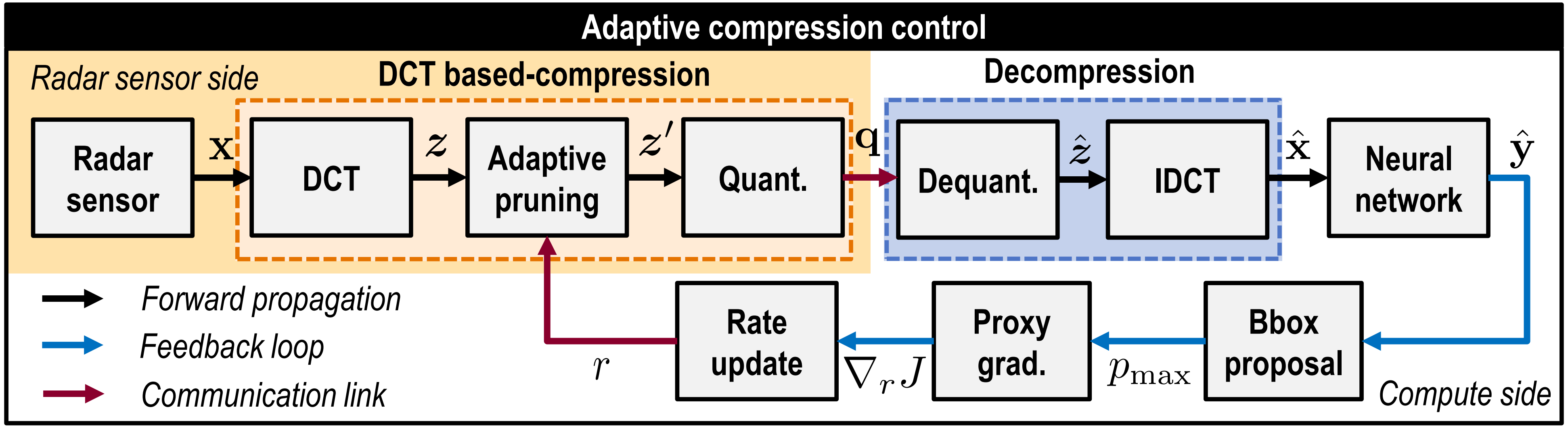
b) Radar signals exhibit extreme sparsity and a heavy-tailed profile (1st/99th percentiles at ±1.74).

→ Necessitating specialized compression beyond standard image-domain codecs.

Table of Contents

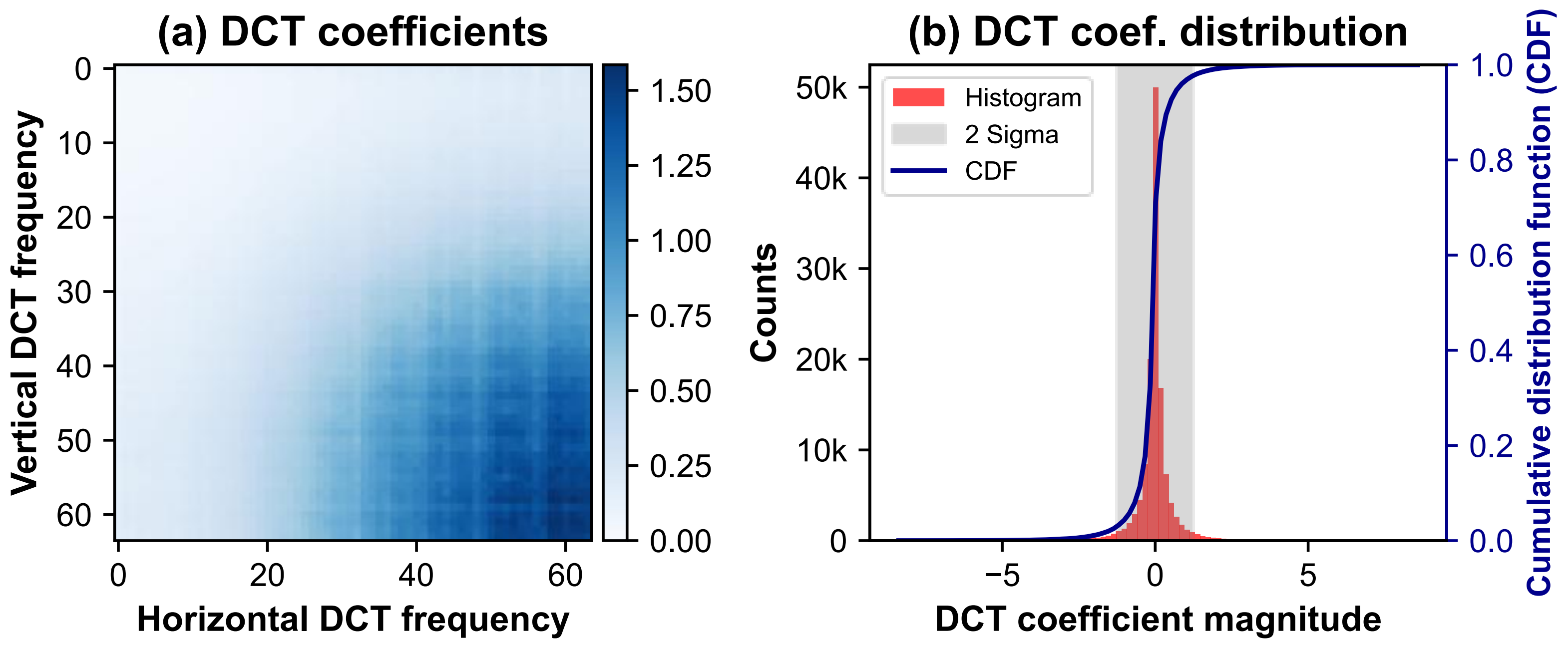
- Motivation
- **Proposed methods**
  - **Spectral pruning**
  - **Adaptive rate feedback**
- Experimental Results
  - Qualitative results
  - Rate-accuracy tradeoff
  - Compression performance against baseline work
- Conclusion

AdaRadar: Adaptive Compression Rate Control



- Our proposed method introduces a **feedback loop** in which the proxy gradient is computed from the detection outputs to update the compression ratio adaptively.
- This avoids the need for **backpropagation** through the communication channel.
- It involves DCT-based **compression** on the radar sensor-side and decompression on the compute side.

Propose Methods: (1) Spectral Compression

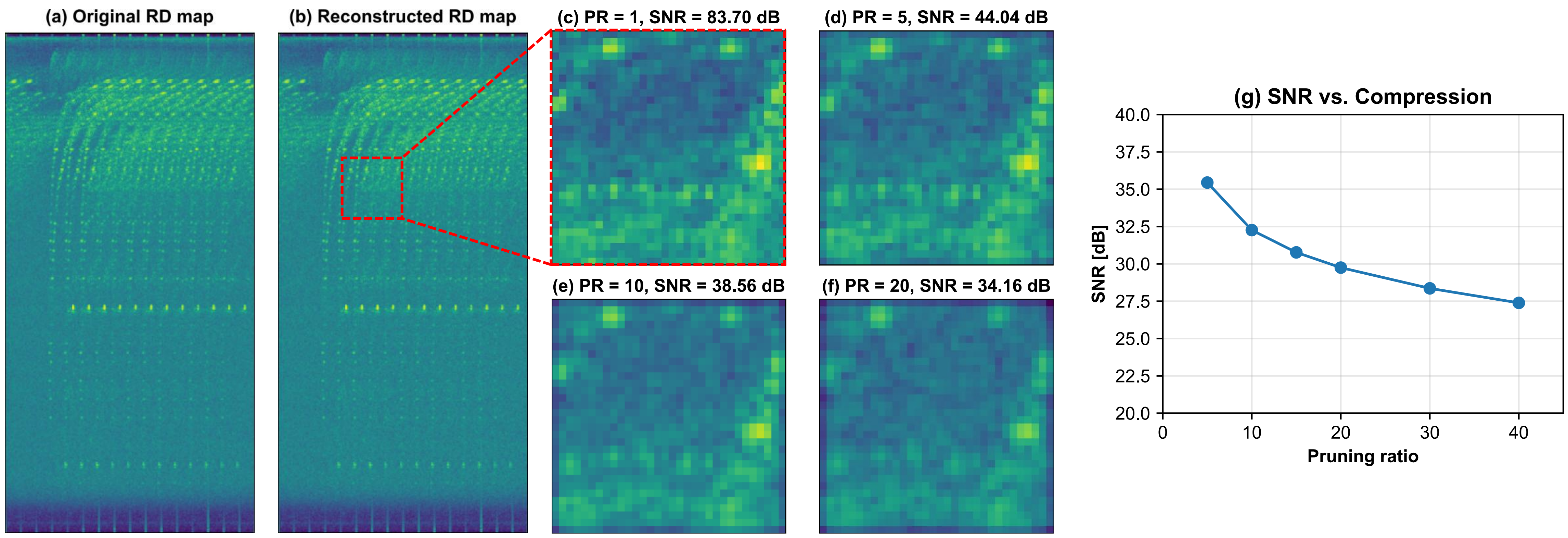


a) The DCT coefficient **magnitudes** are clustered in the high-frequency bins.

b) Their histogram is sharply peaked, highlighting strong **sparsity** and clear opportunities for compression.

→ Adaptive compression includes **spectral pruning** + **symmetric quantization**

Propose Methods: (1) Spectral Compression



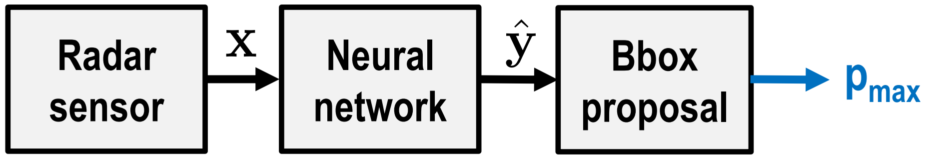
(a) Original and (b) reconstructed **RD map** for a single channel

(c)-(f) Magnified **64 x 64 patches** at pruning ratios of {1,5,10,20}, respectively.

(g) **SNR** vs. bit rate trade-off

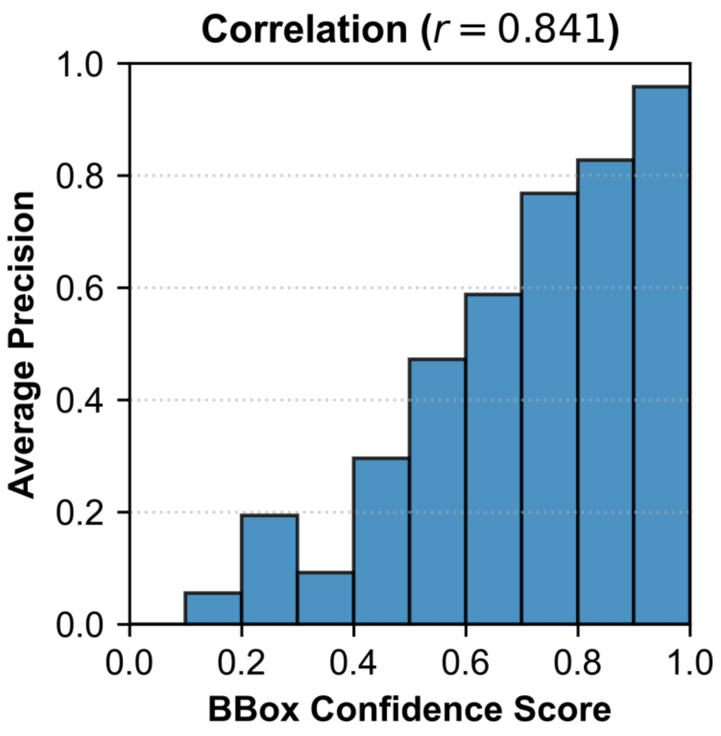
Propose Methods: (2) Adaptive Feedback

Processing chain for radar-based object detection



Bounding boxes (Bbox) in object detection:

Associated Bbox probabilities



- **Labels** are *not* available at test time to monitor the model performance
- We therefore use the bounding box confidence as the **surrogate objective**
- Bbox confidence shows great correlation with the ground truth label (if it were to exist)

Propose Methods: (2) Adaptive Feedback

Task objective:

$$\max_{\{r_t\}_{t=1}^{T}} \mathbb{E}\big[J_t(r_t)\big], \quad J_t(r_t) = h(\mathbf{x}_t, r_t) \; - \; \lambda \cdot B(r_t)$$

hyper-param to control accuracy-bandwidth trade-off

Instantaneous bit-rate

Gradient w.r.t bit-rate (*r*):

$$\nabla_r J = \nabla_r h(\mathbf{x}, r) \; - \; \lambda \cdot B'(r)$$

Proxy gradient:

$$\hat{\nabla}_r h(\mathbf{x}, r) \approx \frac{\overbrace{h(\mathbf{x}, r)}^{\text{Pass \#1}} - \overbrace{h(\mathbf{x}, r-\epsilon)}^{\text{Pass \#2}}}{\epsilon} = \frac{p - p^-}{\epsilon}$$

Algorithm 1: Adaptive rate control

Input: Radar sequence $(\{\mathbf{x}_t\}_{t=1}^{T})$, weight $(\boldsymbol{\Theta})$, initial pruning ratio (r_1), learning rate (η), perturbation (ϵ)

Output: Detection map $(\{\hat{\mathbf{y}}_t\}_{t=1}^{T})$

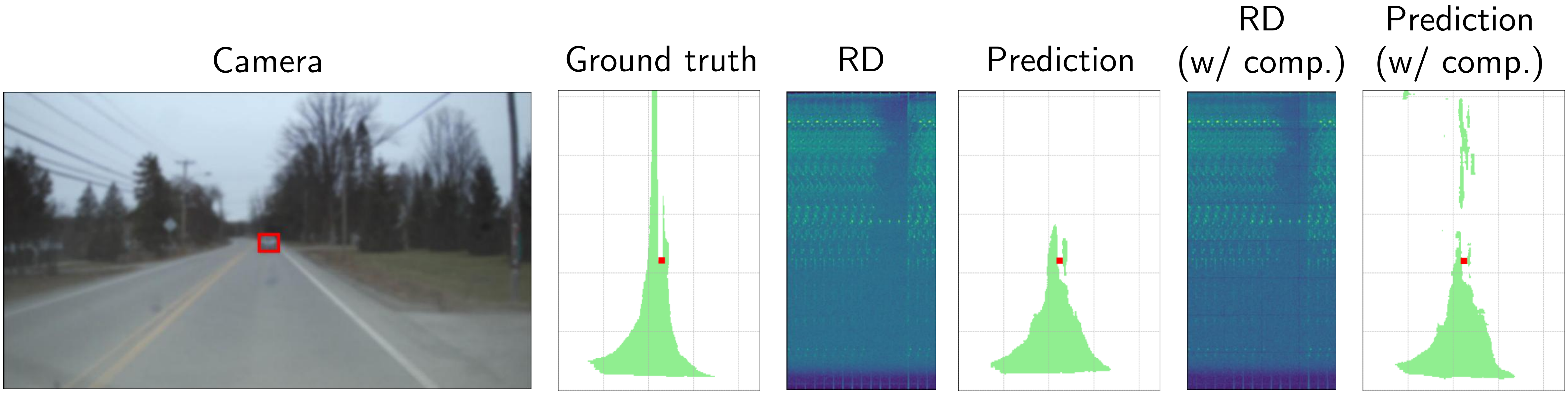
for $t := 1$ *to* T **do**
- $\mathbf{q}_t, \mathbf{Q}_t \leftarrow \text{Compression}(\mathbf{x}_t, r_t)$; `// Transfer` $\{\mathbf{q}, \mathbf{Q}\}$ `once`
- $\hat{\mathbf{x}}_t \leftarrow \text{Decompression}(\mathbf{q}_t, \mathbf{Q}_t)$;
- $\hat{\mathbf{y}}_t \leftarrow f_{\boldsymbol{\Theta}}(\hat{\mathbf{x}}_t)$; ← Network inference pass #1
- $\{(b_k, p_k)\}_{k=1}^{K} = \text{Propose}(\hat{\mathbf{y}}_t)$;
- **if** *adaptive* **then**
  - $\mathbf{q}_t^- \leftarrow \text{Pruning}(\mathbf{q}_t, \epsilon)$;
  - $\hat{\mathbf{x}}_t^- \leftarrow \text{Decompression}(\mathbf{q}_t^-, \mathbf{Q}_t)$;
  - $\hat{\mathbf{y}}_t^- \leftarrow f_{\boldsymbol{\Theta}}(\hat{\mathbf{x}}_t^-)$; ← Network inference pass #2
  - **if** $k > 0$ **then**
    - $r_{t+1} \leftarrow r_t - \eta \hat{\nabla}_{r_t} J$;

- Task objective: *h* encapsulates compression, decompression, forward inference, and Bbox proposal
- Algorithm for adaptive rate control - compute the gradient with only two **forward** passes

Table of Contents

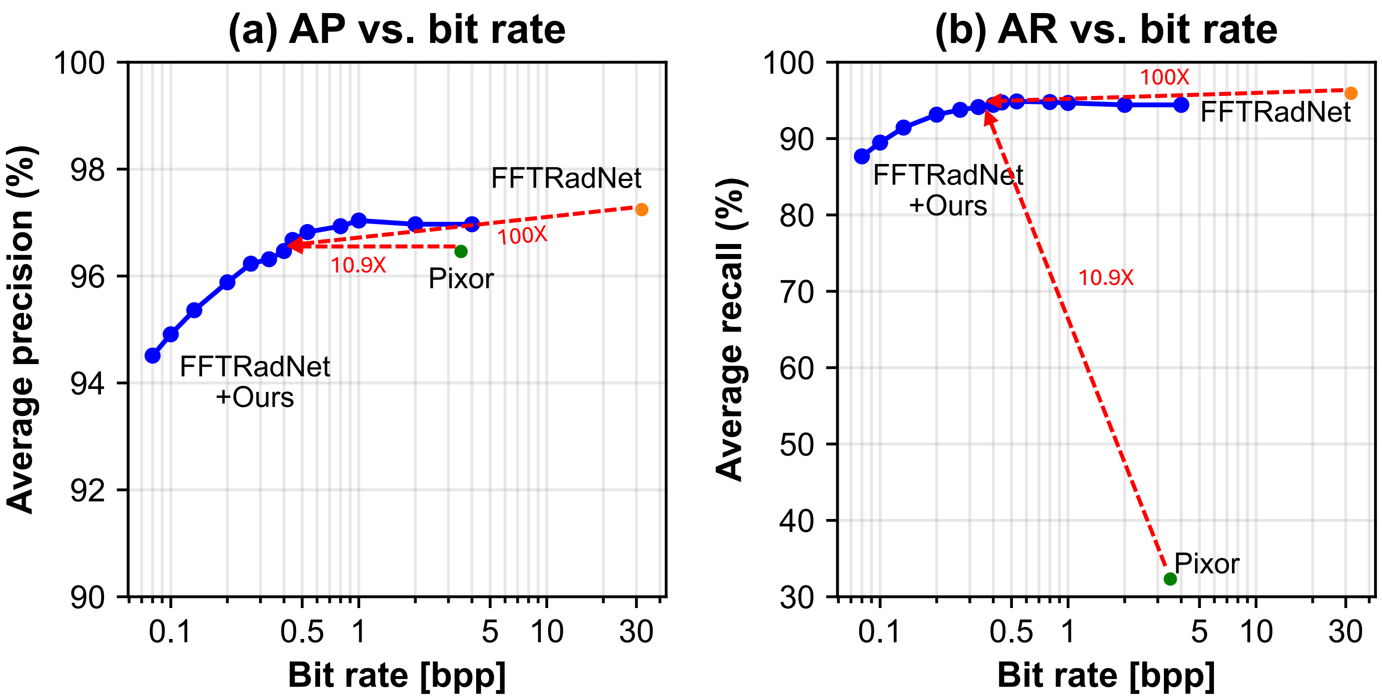
- Motivation
- Proposed methods
  - Spectral pruning
  - Adaptive rate feedback
- **Experimental Results**
  - **Qualitative results**
  - **Rate-accuracy tradeoff**
  - **Compression performance against baseline work**
- Conclusion

Qualitative Results



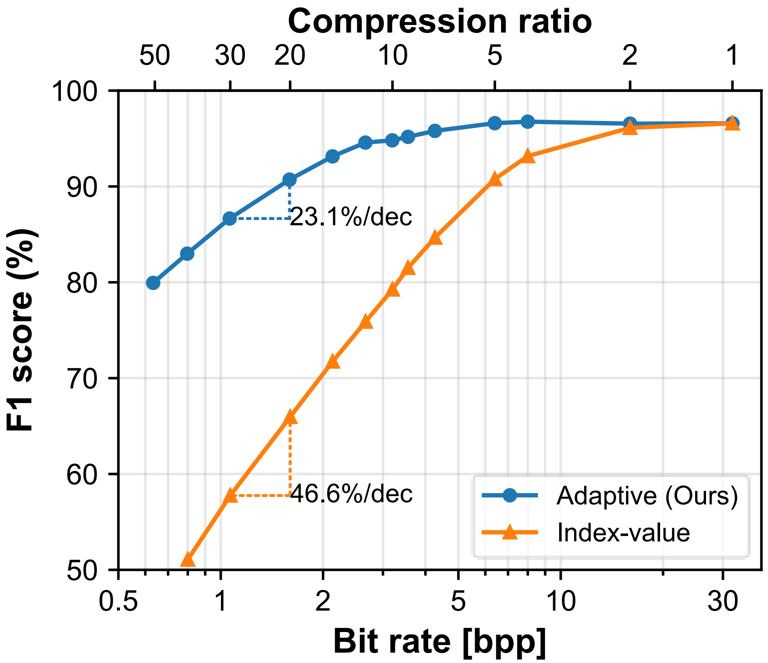
- **Camera images** provide contextual reference for the scene (not fed into network)
- The network predicts both **detection** and **segmentation** outputs.
- Compressed RD leads to **improved performance** compared to the uncompressed baseline.

Rate-accuracy Tradeoff



- It depicts the obvious **rate-accuracy tradeoff**.
- More importantly, our compression scheme achieves **100x** reduction in the radar feature map size while having a **1%p decrease** in the performance compared to the baseline.
- Against **Pixor**, our method achieves better precision and markedly better recall.

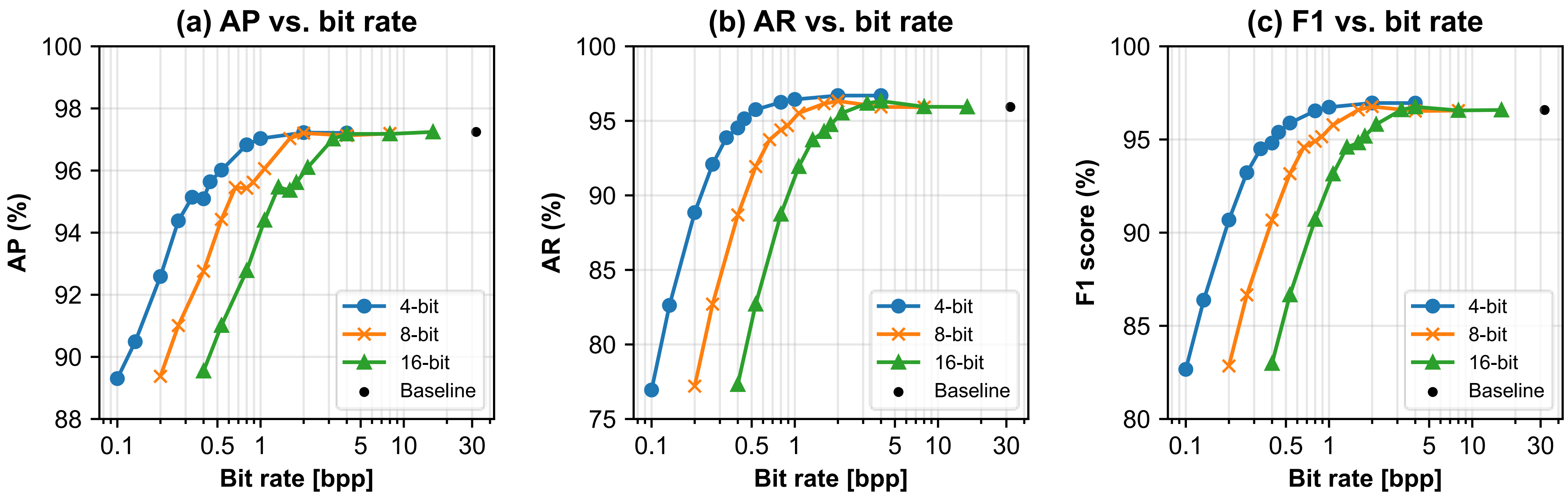
Against Index-value-based Method [NeurIPS'24]



- The **index-value based** compression selects Top-*K* elements in each *M* x *M* block according to the energy in the spatial domain.
- Our method remains **stable** until 5x compression whereas the performance drop-off begins immediately with the index-value pair-based method.
- The roll-off gradient of **23.1%/dec** is much more gentle for the spectral pruning-based compression vs. **46.6%/dec** of the counterpart.

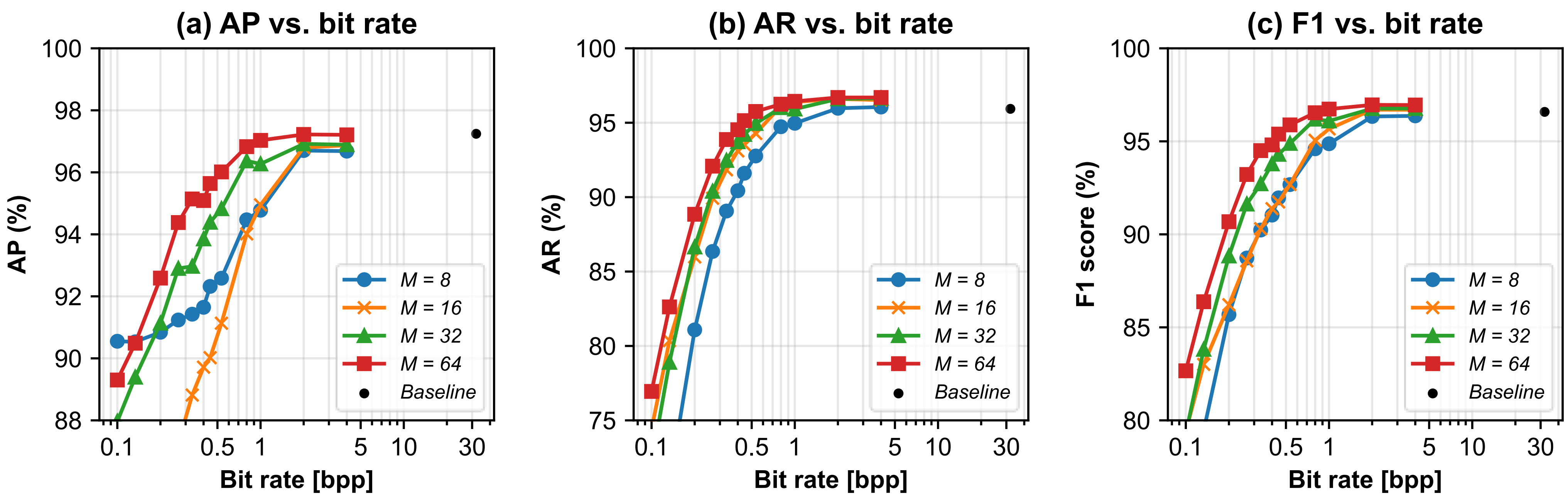
VLSILab
@COLUMBIA UNIV

Effect of Quantization Bit Width on Detection



- We observe that quantization up to 4 bits does not affect the performance compared to that of 8-bit and 16-bit.
- We use the FFTRadNet on the RADIal dataset *without* any fine-tuning with the block length M=64 for all cases.

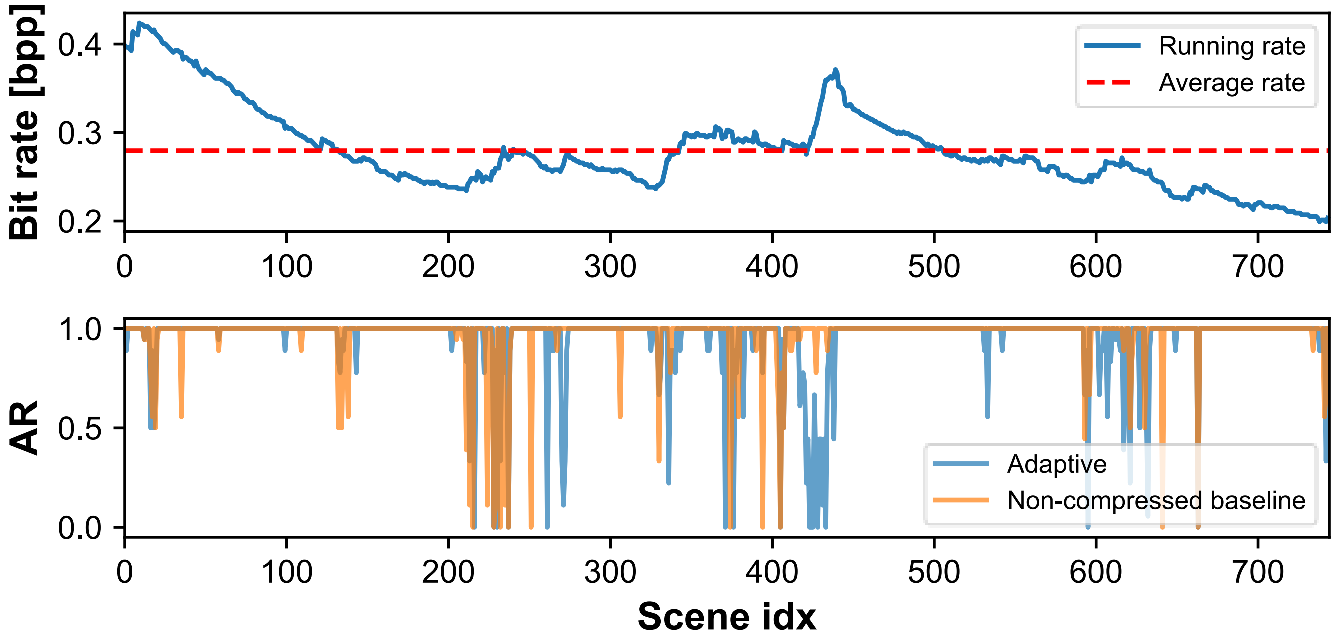
Effect of 2D-DCT Block Length on Detection



- We examine the impact of varying block sizes on compression on the RADIal dataset.
- We show how average precision, recall, and F1-score degrade across block sizes $M \in \{8, 16, 32, 64\}$ with 4-bit quantization.
- Among these, M=64 yields the best trade-off between performance and overhead from the scaling factor.

VLSILab @COLUMBIA UNIV

Online Compression



- We **visualize** the online compression with bit rate and detection metric.
- Here, it is expected that the controller decreases the pruning ratio to **compensate** for the AR drop.
- It achieves an average bit rate of **0.279 bpp** with 8-bit quantization, yielding a **115x** compression with AR of **93.91%**.

Compression Results – RADIal Dataset

RADIal [CVPR'22]
Task: detection

Method	Bit	Prune ratio	Bit rate [bpp] ↓ (Comp. ratio ↑)	Detection P (%) ↑	Detection R (%) ↑	Detection F_1 (%) ↑	Segment. mIoU (%) ↑
Baseline [2]	32	-	32 (1×)	97.24	95.93	96.58	75.97
+Index-value [11]	32	12×	2.67 (12×)	**97.55**	62.12	75.91	49.86
+Adaptive (Ours)	4	12.57×	**0.32** (**101×**)	96.25	**94.04**	**95.13**	**79.34**

- An average pruning ratio of **12.57x** combined with **8x** gain from quantization, yields a bit rate of **0.32 bits per pixel** (bpp), corresponding to a **101x** compression ratio.
- On the segmentation task, ours improves the performance by **4%**.
- Our method markedly **surpasses** index-value based method.

VLSILab @COLUMBIA UNIV

Compression Results – CARRADA, Radatron

CARRADA [ICPR'20]
Task: segmentation

Method	Bit	PR	BR [bpp] ↓	CR ↑	mIoU (%) ↑	mDice (%) ↑
Baseline [41]	32	-	32	1×	55.25	67.13
+Index-value [11]	32	29×	1.10	29×	38.96	46.90
+Adaptive (Ours)	8	29.28×	**0.27**	**117×**	**54.03**	**65.87**

Radatron [ECCV'22]
Task: detection

Method	Bit	PR	BR [bpp] ↓	CR ↑	mAP (%) ↑	AP_{50} (%) ↑	AP_{75} (%) ↑
Baseline [26]	32	-	32	1×	46.07	83.60	44.16
+Index-value [11]	32	7.5×	4.27	7.5×	45.72	80.44	47.54
+Adaptive (Ours)	8	7.5×	**1.07**	**30×**	**48.46**	**83.69**	**49.07**

CARRADA

- It allows even greater compression ratio of **117x** while maintaining the target metric close to the non-compressed baseline with **1%p drop**.

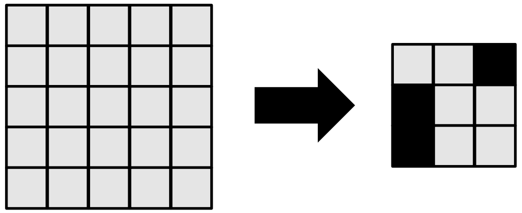
Radatron

- A range-azimuth heatmap is compressible by 30x with our compression method.
- Even at this compression rate, ours actually provides better performance than the non-compressed baseline -- up to **5%** improvement in AP_{75}.

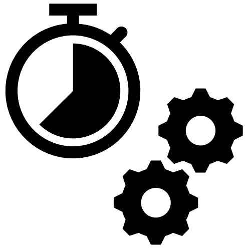
VLSILab @COLUMBIA UNIV

Table of Contents

Conclusion



Spectral feature compression



Online rate adaptation

- We present **AdaRadar**, a rate-adaptive spectral-domain compression framework for radar-based object detection.

- Pruning and quantizing DCT coefficients shrink the range–Doppler cube by >100x, easing the sensor-to-compute bandwidth bottleneck while preserving accuracy at the lowest bit rate target.

- A lightweight test-time controller finely adjusts the compression ratio on the fly, making a judicious tradeoff between bandwidth and performance in real time.

Thank you for listening to our work on AdaRadar:
Rate Adaptive Spectral Compression for Radar-based Perception

Project page



Jinho Park[1] **Se Young Chun**[2] **Mingoo Seok**[1]

[1]Columbia University
[2]Seoul National University
jp4327@columbia.edu, sychun@snu.ac.kr, ms4415@columbia.edu

Acknowledgements:
This work was supported in part by COGNISENSE, one of seven centers in JUMP 2.0, a Semiconductor Research Corporation (SRC) program sponsored by DARPA. The work of SY Chun was supported by IITP grants funded by the Korea government(MSIT) [No.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University) / No.RS-2025-02314125, Effective Human-Machine Teaming With Multimodal Hazy Oracle Models].

CVPR JUNE 3-7, 2026 DENVER COLORADO

VLSILab @COLUMBIA UNIV